EFFECTIVE SYMBOLIC DYNAMICS, RANDOM POINTS, STATISTICAL BEHAVIOR, COMPLEXITY AND ENTROPY

STEFANO GALATOLO, MATHIEU HOYRUP, AND CRISTÓBAL ROJAS

Abstract. We consider the dynamical behavior of Martin-Löf random points in dynamical systems over metric spaces with a computable dynamics and a computable invariant measure. We use computable partitions to define a sort of effective symbolic model for the dynamics. Through this construction we prove that such points have typical statistical behavior (the behavior which is typical in the Birkhoff ergodic theorem) and are recurrent. We introduce and compare some notions of complexity for orbits in dynamical systems and prove: (i) that the complexity of the orbits of random points equals the Kolmogorov-Sinaï entropy of the system, (ii) that the supremum of the complexity of orbits equals the topological entropy.

Contents

PARTLY SUPPORTED BY ANR GRANT 05 2452 260 OX

1. Introduction

The randomness of a particular outcome is always relative to some statistical test. The notion of algorithmic randomness, defined by Martin-Löf in 1966, is an attempt to have an "absolute" notion of randomness. This absoluteness is actually relative to all "effective" statistical tests, and lies on the hypothesis that this class of tests is sufficiently wide.

Martin-Löf's original definition was given for infinite symbolic sequences. With this notion each single random sequence behaves as a generic sequence of the probability space for each effective statistical test. In this way many probabilistic theorems having almost everywhere statements can be translated to statements which hold *for each* random sequence. As an example we cite the fact that in each infinite string of 0's and 1's which is random for the uniform measure, all the digits appear with the same limit frequency. This is a particular case, related to the strong law of large numbers (or Birkhoff ergodic theorem). A general statement of this kind was given by V'yugin (Birkhoff ergodic theorem for individual random sequences, see [V'y97] and lemma 3.2.1 below).

Recently the notion of Martin-Löf randomness was generalized to computable metric spaces endowed with a measure ([Gác05, HR07]). Computable metric spaces are separable metric spaces where the distance can be in some sense effectively computed (see section 2.4). In those spaces, it is also possible to define "computable" functions, which are functions whose behavior is in some sense given by an algorithm, and "computable" measures (there is an algorithm to calculate the measure of nice sets). The space of infinite symbolic sequences, the real line or euclidean spaces, are examples of metric spaces which become computable in a very natural way.

A particularly interesting class of general stationary stochastic processes is constituted by those generated by a measure-preserving map on a metric space, these are the objects studied by ergodic theory. In this paper we consider systems of the type (X, T, μ), where X is a computable metric space, μ a computable probability measure and T a computable endomorphism. The above considered symbolic shifts on spaces of infinite sequences which preserve a computable measure are systems of this kind.

In the classical ergodic theory, a powerful technique (symbolic dynamics) allows to associate to a general system as above (X, T, μ) a shift on a space of infinite strings having similar statistical properties. In section 3 we use the algorithmic features of computable metric spaces to define computable measurable partitions and construct effective symbolic models for the dynamics. In this models *random points are associated to random infinite strings*. This tool allows to generalize theorems which are proved in the symbolic setting to the more general setting of maps and metric spaces. For example the above cited V'yugin theorem becomes a Birkhoff theorem for random points. On this line, we also prove a Poincaré recurrence theorem for random points. Those statements (see thm.3.2.1 and prop. 3.2.1) can be summarized as

Theorem. *Let (X, μ) be a computable probability space. If x is μ-random, then it is recurrent with respect to* every *measure preserving endomorphism T on (X, μ).*

Moreover, each μ-random point x is typical for every *ergodic endomorphism T, i.e.*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) = \int f d\mu \tag{1}$$

for every continuous bounded real-valued f.

In the remaining part of the paper these tools are also used to prove relations between various definitions of orbit complexity and entropy of the systems.

In [Bru83], Brudno defined a notion of algorithmic complexity $\overline{\mathcal{K}}(x,T)$ for the orbits of a dynamical system on a compact space. It is a measure of the information rate which is necessary to describe the behavior of the orbit of x. In this point-wise definition the information is measured by the Kolmogorov information content. Later, White ([Whi93]) also introduced a slightly different version $\underline{\mathcal{K}}(x,T)$. Brudno then proved the following results, later improved by White:

Theorem (Brudno, White)**.** *Let X be a compact topological space and $T : X \to X$ a continuous map.*

1. *For any ergodic probability measure μ the equality $\underline{\mathcal{K}}(x,T) = \overline{\mathcal{K}}(x,T) = h_\mu(T)$ holds for μ-almost all $x \in X$,*
2. *For all $x \in X$, $\overline{\mathcal{K}}(x,T) \leq h(T)$.*

where $h_\mu(T)$ is the Kolmogorov-Sinaï entropy of (X,T) with respect to μ and $h(T)$ is the topological entropy of (X,T). This result seems miraculous as no computability assumption is required on the space or on the transformation T. Actually, this miracle lies in the compactness of the space, which makes it finite when observations are made with finite precision (open covers of the space can be reduced to *finite* open covers). Indeed, when the space is not compact, it is possible to construct systems for which the algorithmic complexity of orbits is correlated in no way to their dynamical complexity. In [Gal00], Brudno's definition was generalized to non-compact computable metric spaces. This definition coincides with Brudno's one in the compact case and will be given in section 4.5.

The above definitions of orbit complexities follow a topological approach. We show that the measure-theoretic setting also provides a natural notion of orbit complexity $\mathcal{K}_\mu(x,T)$ defined by computable partitions. This kind of orbit complexity will be defined almost everywhere and in particular at each μ-random point. For this notion the first result in Brudno and White's theorem comes easily. We go further in showing:

Theorem (4.4.2)**.** *Let T be an ergodic endomorphism of the computable probability space (X,μ),*

$$\mathcal{K}_\mu(x,T) = h_\mu(T) \quad \textit{for all } \mu\textit{-random point } x.$$

We then prove that the two notions of orbit complexity coincide on Martin-Löf random points:

Theorem (5.0.1)**.** *Let T be an ergodic endomorphism of the computable probability space (X,μ), where X is compact,*

$$\mathcal{K}_\mu(x,T) = \overline{\mathcal{K}}(x,T) \quad \textit{for all } \mu\textit{-random point } x.$$

In the topological context, we then consider $\overline{\mathcal{K}}(x,T)$ and strengthen the second part of Brudno's theorem, showing:

Theorem (6.3.1)**.** *Let T be a computable map on a compact computable metric space X,*

$$\sup_{x\in X}\underline{\mathcal{K}}(x,T)=\sup_{x\in X}\overline{\mathcal{K}}(x,T)=h(T)$$

Remark that this was already implied by Brudno's theorem, using the variational principle: $h(T)=\sup\{h_\mu(T):\mu \text{ is } T\text{-invariant}\}$. Nevertheless, our proof uses purely topological and algorithmic arguments and no measure. In particular, it does not use the variational principle, and can be thought as an alternative proof of it.

Many of these statements require that the dynamics and the invariant measure are computable.

The first assumption can be easily checked on concrete systems if the dynamics is given by a map which is effectively defined.

The second is more delicate: it is well known that given a map on a metric space, there can be a continuous (even infinite dimensional) space of probability measures which are invariant for the map, and many of them will be non computable. An important part of the theory of dynamical systems is devoted to selecting measures which are particularly meaningful. From this point of view, an important class of these measures is the class of SRB invariant measures, which are measures being in some sense the "physically meaningful ones"(for a survey on this topic see [You02]). It can be proved (see [GHR07b] and [GHR07a] and their references e.g.) that in several classes of dynamical systems where SRB measures are proved to exist, these measures are also computable from our formal point of view, hence providing several classes of nontrivial concrete examples where our results can be applied.

2. Preliminaries

2.1. **Partial recursive functions.** The notion of algorithm working on integers has been formalized independently by Markov, Church, Turing among others. Each constructed model defines a set of partial (not defined everywhere) integer functions which can be computed by some *effective* mechanical or algorithmic (w.r.t. the model) procedure. Later, it has been proved that all this models define the same class of functions, namely: the set of *partial recursive functions*. This fact supports a working hypothesis known as Church's Thesis, which states that every (intuitively formalizable) algorithm is a partial recursive function. It gives the connection between the informal notion of algorithm and the formal definition of recursive function.

Let us say then that a ***recursive function*** is a function (on integers) that can be computed in some *effective* or *algorithmic* way. For formal definitions see for example, [Rog87]. With this intuitive description it is more or less clear that there exists an effective procedure to enumerate the class of all partial recursive functions, associating to each of them its ***Gödel number***, which is the number of the program computing it. Hence there exists a universal recursive function $\varphi_u:\mathbb{N}\to\mathbb{N}$ satisfying for all $e,n\in\mathbb{N}$, $\varphi_u(\langle e,n\rangle)=\varphi_e(n)$ where e is the gödel number of φ_e and $\langle,\rangle:\mathbb{N}^2\to\mathbb{N}$ is some recursive bijection. In classical recursion theory, a set of natural numbers is called ***recursively enumerable*** (***r.e*** for short) if it is the range of some partial recursive function. That is if there exists an algorithm listing the set. We denote by E_e the r.e set associated to φ_e. Namely $E_e=rang(\varphi_e)=\{\varphi_u(\langle e,n\rangle):n\in\mathbb{N}\}$.

2.2. **Algorithms on finite objects.** Strictly speaking, algorithms only work on integers. However, when the objects of some class have been identified with integers, it makes sense to speak about algorithms acting on these objects.

Definition 2.2.1. A ***Numbered Set*** $\mathcal{O}$ is a countable set together with a surjection $\nu_{\mathcal{O}} : \mathbb{N} \to \mathcal{O}$ called the ***numbering***. We write o_n for $\nu(n)$ and call n the ***name*** of o_n.

Of course, the potential of algorithms depends on the choice of the numbering, since it determines to what extent an object can be algorithmically recovered from its name. If the objects of some collection can be characterized by a finite number of integers, then the collection is a numbered set since its objects can be indexed using standard recursive bijections from $\mathbb{N}^*$ to $\mathbb{N}$ and from $\mathbb{N}^k$ to $\mathbb{N}$, which we will both denote $\langle . \rangle$.

Examples.

(1) $\mathbb{Q}$, with some standard numbering $\nu_{\mathbb{Q}}$ is a numbered set.
(2) The set of partial recursive functions $\mathcal{R} = \{\varphi_e : e \in \mathbb{N}\}$ is a numbered set, Gödel numbers being the names.
(3) The collection $\{E_e = rang(\varphi_e) : e \in \mathbb{N}\}$ of all r.e subsets of $\mathbb{N}$ is a numbered set.

Definition 2.2.2. Let $\mathcal{O}$ be a numbered set. To any recursive function $\varphi : \mathbb{N} \to \mathbb{N}$ we associate an ***algorithm*** $\mathcal{A}_\varphi : \mathbb{N} \to \mathcal{O}$ defined by $\mathcal{A}_\varphi(i) = o_{\varphi(i)}$.

Given a total algorithm $\mathcal{A} : \mathbb{N} \to \mathcal{O}$ (i.e $\mathcal{A} = \mathcal{A}_\varphi$ for some total φ), we say that the sequence of finite objects $(\mathcal{A}(n))_{n\in\mathbb{N}}$ is ***enumerated by*** $\mathcal{A}$, or that $\mathcal{A}$ is an ***algorithmic enumeration*** of this sequence.

2.3. **Computability over the reals.**

Definition 2.3.1. Let x be a real number. We say that:

- x is ***lower semi-computable*** if the set $E := \{i \in \mathbb{N} : q_i < x\}$ is r.e,
- x is ***upper semi-computable*** if the set $E := \{i \in \mathbb{N} : q_i > x\}$ is r.e,
- x is ***computable*** if it is lower and upper semi-computable.

Equivalently, a real number is computable if and only if there exists an algorithmic enumeration of a sequence of rational numbers converging exponentially fast to x. That is:

Proposition 2.3.1. *A real number x is computable if and only if there exists an algorithm $\mathcal{A} : \mathbb{N} \to \mathbb{Q}$ such that $|\mathcal{A}(i) - x| < 2^{-i}$, for all i.*

Definition 2.3.2. Let $(x_n)_n$ be a sequence of computable reals. We say that the sequence is ***uniformly*** computable or that x_n is computable ***uniformly in n*** if there exists an algorithm $\mathcal{A} : \mathbb{N} \to \mathbb{Q}$ such that for all n and i it holds $|\mathcal{A}(\langle n, i\rangle) - x_n| < 2^{-i}$.

Uniform sequences of lower (upper) semi-computable reals are defined in the same way.

2.4. **Computable Metric Spaces.**

Definition 2.4.1. A ***computable metric space*** (CMS) is a triple $\mathcal{X} = (X, d, \mathcal{S})$, where

- (X, d) is a separable complete metric space.
- $\mathcal{S} = (s_i)_{i\in\mathbb{N}}$ is a numbered dense subset of X (called ***ideal points***).
- The real numbers $(d(s_i, s_j))_{\langle i,j\rangle\in\mathbb{N}}$ are all computable, uniformly in $\langle i, j\rangle$.

We now recall some important examples of computable metric spaces:

Examples. (1) the Cantor space $(\Sigma^{\mathbb{N}}, d, S)$ with Σ a finite alphabet and d the usual distance. S is the set of ultimately 0-stationary sequences.
(2) $(\mathbb{R}^n, d_{\mathbb{R}^n}, \mathbb{Q}^n)$ with the euclidean metric and the standard numbering of $\mathbb{Q}^n$.
(3) if (X_1, d_1, S_1) and (X_2, d_2, S_2) are two computable metric spaces, the distance $d((x_1, x_1), (y_1, y_2)) = \max(d_1(x_1, y_1), d_2(x_2, y_2))$ on the product space $(X_1 \times X_2, d, S_1 \times S_1)$ makes it a computable metric space.

For further examples we refer to [Wei93].

Let $(X, d, \mathcal{S})$ be a computable metric space. The computable structure of X assures that the whole space can be "reached" using algorithmic means. Since ideal points (the finite objects of $\mathcal{S}$) are dense, they can approximate any x at any finite precision. Then, x itself can be identified to a sequence of ideal points converging to x in an effectively controlled way. Let us say that a sequence of ideal points $(s_{i_n})_n$ is ***fast*** if $d(s_{i_n}, s_{i_{n+1}}) < 2^{-n}$ for all n. As the space is complete, a fast sequence has always a limit x, and $d(s_{i_n}, x) < 2^{-(n-1)}$ for all n.

Definition 2.4.2 (Computable points)**.** A point $x \in X$ is said to be ***computable*** if there exists an algorithm $\mathcal{A} : \mathbb{N} \to \mathcal{S}$ which enumerates a fast sequence whose limit is x.

As for real numbers we can give the notion of uniform sequence

Definition 2.4.3. Let $(x_n)_n$ be a sequence of computable points. We say that the sequence is ***uniformly*** computable or that x_n is computable ***uniformly in n*** if there exists an algorithm $\mathcal{A} : \mathbb{N} \to \mathcal{S}$ such that for all n, the sequence $(\mathcal{A}(\langle n, i \rangle))_i$ is fast and converges to x_n.

The numbered set of ideal points $(s_i)_i$ induces the numbered set of ***ideal balls*** $\mathcal{B} := \{B(s_i, q_j) : s_i \in S, q_j \in \mathbb{Q}_{>0}\}$. We denote by $B_{\langle i,j \rangle}$ the ideal ball $B(s_i, q_j)$.

Computability can then be extended from the numbered set $\mathcal{B}$ to the space of open subsets of X: such an open subset $U \subseteq X$ can be identified to a collection of ideal balls whose union is U.

Definition 2.4.4 (R.e open sets)**.** We say that the set $U \subset X$ is ***r.e open*** if there is some r.e set $E \subset \mathbb{N}$ such that $U = \cup_{i \in E} B_i$.

Remark 2.4.1. Let U be a r.e open set. It is easy to see that there is an algorithm to ***semi-decide*** weather some ideal point belongs to U. That is, the algorithm will halt on input i iff $s_i \in U$. This notion can be extended to any point x in the following sense: The algorithm sequentially asks (from an external user) for finite approximations of x at required precisions. If $x \in U$ the algorithm will eventually stop and answer "yes", if $x \notin U$ then the algorithm will run and ask forever. For formal definitions we refer to [HR07].

Definition 2.4.5. Let $(U_n)_n$ be a sequence of r.e open sets. We say that the sequence is ***uniformly r.e*** or that U_n is r.e open ***uniformly in n*** if there exists an r.e set $E \subset \mathbb{N}$ such that for all n it holds $U_n = \cup_{i \in E_n} B_i$, where $E_n = \{i : \langle n, i \rangle \in E\}$.

Examples. (1) If the sequence $(U_n)_n$ is uniformly r.e then the union $\cup_n U_n$ is a r.e open set.

(2) The universal recursive function φ_u makes the collection of all r.e open sets (denoted $\mathcal{U}$) a sequence uniformly r.e. Indeed, define $E := \{\langle e, \varphi_u(\langle e, n\rangle)\rangle : e, n \in \mathbb{N}\}$. Then $\mathcal{U} = \{U_e : e \in N\}$ where $U_e = \cup_{i\in E_e} B_i$.
(3) The numbered set $\mathcal{U}$ is closed under finite unions and finite intersections. Furthermore, these operations are *effective* in the following sense: there exists recursive functions $\varphi^{\cup}$ and $\varphi^{\cap}$ such that for all $i, j \in \mathbb{N}$, $U_i \cup U_j = U_{\varphi^{\cup}(\langle i,j\rangle)}$ and the same holds for $\varphi^{\cap}$. Equivalently: $U_i \cup U_j$ is r.e open uniformly in $\langle i, j\rangle$. See [HR07].

Definition 2.4.6 (Constructive G_δ-sets)**.** We say that the set $D \subset X$ is a ***constructive G_δ-set*** if it is the intersection of a sequence of uniformly r.e open sets.

Let (X, S_X, d_X) and (Y, S_Y, d_Y) be computable metric spaces with $\mathcal{U}^X$ and $\mathcal{U}^Y$ the corresponding numbered sets of r.e open sets.

Definition 2.4.7 (Computable Functions)**.** A function $T : X \to Y$ is said to be ***computable*** if $T^{-1}(B_n)$ is r.e open uniformly in n.

Remark 2.4.2. We remark that this definition implies that the preimage of a uniform sequence of r.e. open sets is a uniform sequence of r.e. open sets. This could be an alternative definition of computable function.

It follows that computable functions are continuous. Since we will work with functions which are not necessarily continuous everywhere, we shall consider functions which are computable on some subset of X. More precisely, a function T is said to be ***computable on D*** ($D \subset X$) if there is a uniform sequence $(U_n^X)_n$ of r.e open subsets of X such it holds $T^{-1}(B_n) \cap D = U_n^X \cap D$ for the uniform sequence of ideal balls B_n. D is called the ***domain of computability*** of T.
remarks:

- Since ideal balls generate the topology, a function is computable iff $T^{-1}(B_n^Y)$ is r.e open uniformly in n (or the intersection of D with a uniformly r.e open set).
- If T is computable then the images of ideal points can be uniformly computed, that is: $T(s_i^X)$ is a computable point, uniformly in i.
- More generally, if T is computable then there exists an algorithm which computes the image $T(x)$ of any x in the following sense: the user enters some rational ϵ to the algorithm which, after asking finitely many times the user for finite approximations of x, halts outputting a finite approximation of $T(x)$ up to ϵ.
- The distance function $d : X \times X \to \mathbb{R}$ is a computable function.

2.5. **Computable Probability Spaces (CPS).** When X is a computable metric space, the space of probability measures over X, denoted by $\mathcal{M}(X)$, can be endowed with a structure of computable metric space (this will be defined below, for more details, see [Gác05, HR07]). Then a computable measure can be defined as a computable point of $\mathcal{M}(X)$.

Some prerequisites from measure theory: We say that μ_n converge weakly to μ and write $\mu_n \to \mu$ if

$$\mu_n \to \mu \text{ iff } \mu_n f \to \mu f \text{ for all real continuous bounded } f \tag{2}$$

where μf stands for $\int f d\mu$. Let us recall the Portmanteau theorem. We say that a Borel set A is ***μ-continuous*** if $\mu(\partial A) = 0$, where $\partial A = \overline{A} \cap \overline{X \setminus A}$ is the boundary of A.

Theorem 2.5.1 (Portmanteau theorem)**.** *Let μ_n, μ be Borel probability measures on a separable metric space (X, d). The following are equivalent:*

(1) μ_n converges weakly to μ,
(2) $\limsup_n \mu_n(F) \leq \mu(F)$ for all closed sets F,
(3) $\liminf_n \mu_n(G) \geq \mu(G)$ for all open sets G,
(4) $\lim_n \mu_n(A) = \mu(A)$ for all μ-continuity sets A.

This theorem easily implies the following: when (X, d) is a separable metric space, weak convergence can be proved using the following criterion:

Proposition 2.5.1. *Let $\mathcal{A}$ be a countable basis of the topology which is closed under the formation of finite unions. If $\mu_n(A) \to \mu(A)$ for every $A \in \mathcal{A}$, then μ_n converge weakly to μ.*

Let us introduce on $\mathcal{M}(X)$ the structure of a computable metric space. Let us endow $\mathcal{M}(X)$ with the weak topology, which is the topology of weak convergence. As X is separable and complete, so is $\mathcal{M}(X)$. Let $D \subset \mathcal{M}(X)$ be the set of those probability measures that are concentrated in finitely many points of S and assign rational values to them. It can be shown that this is a dense subset ([Bil68]).

We consider the Prokhorov metric ρ on $\mathcal{M}(X)$ defined by:

$$\rho(\mu, \nu) := \inf\{\epsilon \in \mathbb{R}^+ : \mu(A) \leq \nu(A^\epsilon) + \epsilon \text{ for every Borel set } A\}.$$

where $A^\epsilon = \{x : d(x, A) < \epsilon\}$.

This metric induces the weak topology on $\mathcal{M}(X)$. Furthermore, it can be shown that the triple $(\mathcal{M}(X), D, \rho)$ is a computable metric space (see [Gác05], [HR07]).

Definition 2.5.1. A measure μ is computable if there is an algorithmic enumeration of a fast sequence of ideal measures $(\mu_n)_{n\in\mathbb{N}} \subset D$ converging to μ in the Prokhorov metric and hence, in the weak topology.

The following theorem gives a characterization for the computability of measures in terms of the computability of the measure of sets (for a proof see [HR07]):

Theorem 2.5.2. *A measure $\mu \in \mathcal{M}(X)$ is computable if and only if the measure $\mu(B_{i_1} \cup \ldots \cup B_{i_k})$ of finite unions of ideal open balls is lower-semi-computable uniformly in $\langle i_1, \ldots, i_k \rangle$.*

Definition 2.5.2. A ***Computable Probability Space (CPS)*** is a pair $(\mathcal{X}, \mu)$ where $\mathcal{X}$ is a computable metric space and μ is a computable Borel probability measure on X.

Definition 2.5.3. Let $(\mathcal{X}, \mu)$ and $(\mathcal{Y}, \nu)$ be two computable probability spaces. A ***morphism*** from $(\mathcal{X}, \mu)$ to $(\mathcal{Y}, \nu)$ is a measure-preserving function $F : X \to Y$ which is computable on a constructive G_δ-set of μ-measure one.

We recall that F is measure-preserving if $\nu(A) = \mu(F^{-1}(A))$ for every Borel set A. Computable probability structures can be easily transferred:

Proposition 2.5.2. *Let $(\mathcal{X}, \mu)$ be a computable probability space, $\mathcal{Y}$ be a computable metric space and $F : X \to Y$ a function which is computable on a constructive G_δ-set of μ-measure one. Then the induced measure μ_F on Y defined by $\mu_F(A) = \mu(F^{-1}(A))$ is computable and F is a morphism of computable probability spaces.*

2.6. **Algorithmic randomness.** Now we consider a generalization of Martin-Löf tests to computable probability spaces. Let $(\mathcal{X}, \mu)$ be a computable probability space.

Definition 2.6.1. A ***Martin-Löf μ-Test*** is a sequence $(U_n)_{n\in\mathbb{N}}$ of uniformly r.e open sets which satisfy $\mu(U_n) < 2^{-n}$ for all n. Any subset of $\bigcap_n U_n$ is called an ***effective μ-null set***.

Definition 2.6.2. A point $x \in X$ is called ***μ-random*** if x is contained in no effective μ-null set. The set of μ-random points is denoted R_μ.

Note that $\mu(R_\mu) = 1$. The following is the generalization for metric spaces of a classical result in Cantor space due to Martin-Löf. It says that the set of non-random points is not only a null set but an effective null set. For a proof see [HR07].

Theorem 2.6.1. *The union of all effective μ-null sets, denoted by $\mathcal{N}_\mu$, is again an effective μ-null set.*

Thus, there is a single Martin-Löf test (often called *universal*) which tests non-randomness, and $R_\mu = \mathcal{N}_\mu^c$.

We will need the following results, also taken from [HR07].

Lemma 2.6.1. *Every μ-random point is in every r.e open set of full measure.*

Proposition 2.6.1 (Morphisms of CPS preserve randomness)**.** *Let F be a morphism of computable probability spaces $(\mathcal{X}, \mu)$ and $(\mathcal{Y}, \nu)$. Then every μ-random point x is in the domain of computability of F and $F(x)$ is ν-random.*

2.7. **Kolmogorov complexity.** The idea is to define, for a finite object, the minimal amount of algorithmic information from which the object can be recovered. That is, the length of the shortest description (code) of the object. Since this shortest description is supposed to contain all necessary information to reconstruct in an algorithmic way the coded finite object, the Kolmogorov Complexity is also called *Algorithmic Information Content*. For a complete introduction to Kolmogorov complexity we refer to a standard text [LV93].

Let Σ^* and $\Sigma^\mathbb{N}$ be the sets of finite and infinite words (over the finite alphabet Σ) respectively. A word $w \in \Sigma^*$ defines the *cylinder* $[w] \subset \Sigma^\mathbb{N}$ of all possible continuations of w. A set $D = \{w_1, w_2, ...\} \subset \Sigma^*$ defines an open set $[D] = \cup_i [w_i] \subset \Sigma^\mathbb{N}$. D is called prefix-free if no word of D is prefix of another one, that is if the cylinders $[w_i]$ are pairwise disjoint.

Let X be Σ^* or $\mathbb{N}$ or $\mathbb{N}^*$.

Definition 2.7.1. An ***interpreter*** is a partial recursive function $\varphi : \{0,1\}^* \to X$ which has a prefix-free domain.

Definition 2.7.2. Let $I : \{0,1\}^* \to X$ be an interpreter. The ***complexity*** (or ***Information Content***) $K_I(x)$ of $x \in X$ is defined to be

$$K_I(x) = \begin{cases} |p| & \text{if } p \text{ is a shortest input such that } I(p) = x \\ \infty & \text{if there is no } p \text{ such that } I(p) = x \end{cases}$$

It turns out that there exists an algorithmic enumeration of all interpreters, which entails the existence of a universal interpreter U which is asymptotically optimal in the sense that the *invariance theorem* holds:

Theorem 2.7.1 (Invariance theorem)**.** *For all interpreter I there exists $c_I \in \mathbb{N}$ such that for all $x \in X$ we have $K_U(w) \leq K_I(x) + c_I$.*

We fix a universal interpreter U and we let $K(x) = K_U(x)$.

2.7.1. *Estimates.* Let us recall some simple estimates of complexity. Let f, g be real-valued functions. We say that g ***additively dominates*** f and write $f \overset{+}{<} g$ if there is a constant c such that $f \leq g + c$. As codes are always *binary* words, we use base-2 logarithms, which we denote by log. We define $J(x) = x + 2\log(x+1)$ for $x \geq 0$.

For $n \in \mathbb{N}$, $K(n) \overset{+}{<} J(\log n)$. For $n_1, \ldots, n_k \in \mathbb{N}$, $K(n_1, \ldots, n_k) \overset{+}{<} K(n_1) + \ldots + K(n_k)$. The following property is a version of a result attributed to Kolmogorov, stated in terms of prefix complexity instead of plain complexity.

Proposition 2.7.1. *Let $E \subseteq \mathbb{N} \times X$ be a r.e. set such that $E_n = \{x : (n, x) \in E\}$ is finite for all n. Then for (s, n) with $s \in E_n$,*

$$K(s) \overset{+}{<} J(\log |E_n|) + K(n)$$

Proposition 2.7.2. *Let μ be a computable measure on $\Sigma^{\mathbb{N}}$. For all $w \in \Sigma^*$,*

$$K(w) \overset{+}{<} -\log \mu([w]) + K(|w|)$$

Theorem 2.7.2 (Coding theorem)**.** *Let $P : X \to \overline{\mathbb{R}}^+$ be a lower semi-computable function such that $\sum_x P(x) \leq 1$. Then $K(x) \overset{+}{<} -\log P(x)$, i.e. there is a constant c such that $K(x) \leq -\log P(x) + c$ for all $x \in X$.*

Moreover, $\sum_x 2^{-K(x)} \leq 1$ as it is the Lebesgue measure of the domain of the universal interpreter U. There is a relation between Kolmogorov complexity and randomness, initial segments of random infinite strings being maximally complex.

Theorem 2.7.3 (Chaitin, Levin)**.** *Let μ be a computable measure. Then $\omega \in \Sigma^{\mathbb{N}}$ is a μ-random sequence if and only if $\exists m\ \forall n\ K(\omega_{1:n}) \geq -\log \mu[\omega_{1:n}] - m$.*

The minimal such m, defined by $d_\mu(\omega) := \sup_n\{-\log \mu[\omega_{1:n}] - K(\omega_{1:n})\}$ and called the ***randomness deficiency*** of ω w.r.t μ, is not only finite almost everywhere: it has finite mean, that is $\int d_\mu(\omega) d\mu \leq 1$. For a proof see [LV93].

3. Effective symbolic dynamics and statistics of random points

Let $(\mathcal{X}, \mu)$ be a computable probability space and let R_μ be the set of random points. The aim of this section is to study the set R_μ from a dynamical point of view. That is, we will put a dynamic T on $(\mathcal{X}, \mu)$ (an endomorphism of computable probability space), and look at the abilities of random points (which are *a priori* independent of T) to describe the statistical properties of T.

We recall that a Borel set A is called T***-invariant*** if $T^{-1}(A) = A (\text{mod } 0)$ and that the transformation T is said to be ***ergodic*** if every T-invariant set has measure 0 or 1.

3.1. **Symbolic dynamics of random points.** Let T be an endomorphism of the (Borel) probability space (X, μ). In the classical construction, one considers access to the system given by a finite measurable partition, that is a finite collection of pairwise disjoint Borel sets $\mathcal{P} = \{p_1, \ldots, p_k\}$ such that $\mu(\cup_i p_i) = 1$. Then, to (X, μ, T) is associated a *symbolic dynamical system* $(X_{\mathcal{P}}, \sigma)$ (called the symbolic

model of $(X, T, \mathcal{P})$). The set $X_{\mathcal{P}}$ is a subset of $\{1, 2, \ldots, k\}^{\mathbb{N}}$. To a point $x \in X$ corresponds an infinite sequence $\omega = (\omega_i)_{i\in\mathbb{N}} = \phi_{\mathcal{P}}(x)$ defined by:

$$\phi_{\mathcal{P}}(x) = \omega \Leftrightarrow \forall j \in \mathbb{N}, T^j(x) \in p_{\omega_j}$$

The transformation $\sigma : X_{\mathcal{P}} \to X_{\mathcal{P}}$ is *the shift* defined by $\sigma((\omega_i)_{i\in\mathbb{N}}) = (\omega_{i+1})_{i\in\mathbb{N}}$.

As $\mathcal{P}$ is a measurable partition, the map $\phi_{\mathcal{P}}$ is measurable and then the measure μ induces the measure $\mu_{\mathcal{P}}$ (on the associated symbolic model) defined by $\mu_{\mathcal{P}}(B) = \mu(\phi_{\mathcal{P}}^{-1}(B))$ for all measurable $B \subset X_{\mathcal{P}}$.

The requirement of $\phi_{\mathcal{P}}$ being measurable makes the symbolic model appropriate from the measure-theoretic view point, but is not enough to have a symbolic model compatible with the computational approach:

Definition 3.1.1. Let T be an endomorphism of the computable probability space $(\mathcal{X}, \mu)$ and $\mathcal{P} = \{p_1 \ldots, p_k\}$ a finite measurable partition. The associated symbolic model $(X_{\mathcal{P}}, \mu_{\mathcal{P}}, \sigma)$ is said to be ***an effective symbolic model*** if the map $\phi_{\mathcal{P}} : X \to \{1, \ldots, k\}^{\mathbb{N}}$ is a morphism of CPS (here the space $\{1, \ldots, k\}^{\mathbb{N}}$ is endowed with the standard computable structure).

The sets p_i are called the ***atoms*** of $\mathcal{P}$ and we denote by $\mathcal{P}(x)$ the atom containing x (if there is one). Observe that $\phi_{\mathcal{P}}$ is computable on its domain only if the atoms are open r.e sets (in the domain):

Definition 3.1.2 (Computable partitions)**.** A measurable partition $\mathcal{P}$ is said to be a ***computable partition*** if its atoms are r.e open sets.

Conversely:

Theorem 3.1.1. *Let T be an endomorphism of the CPS (X, μ) and $\mathcal{P} = \{p_1 \ldots, p_k\}$ a finite computable partition. Then the associated symbolic model is effective.*

Proof. Let D be the domain of computability of T (it is a full-measure constructive G_δ). Define the set

$$X^{\mathcal{P}} = D \cap \bigcap_{n\in\mathbb{N}} T^{-n}(p_1 \cup \ldots \cup p_k)$$

$X^{\mathcal{P}}$ is a full-measure constructive G_δ-set: indeed, as $p_1 \cup \ldots \cup p_k$ is r.e. and T is computable on D there are uniformly r.e. open sets U_n such that $D \cap T^{-n}(p_1 \cup \ldots \cup p_k) = D \cap U_n$, so $X^{\mathcal{P}} = D \cap \bigcap_n U_n$. As T is measure-preserving, all U_n have measure one.

Now, $X^{\mathcal{P}} \cap \phi_{\mathcal{P}}^{-1}[i_0, \ldots, i_n] = X^{\mathcal{P}} \cap p_{i_0} \cap T^{-1} p_{i_1} \cap \ldots \cap T^{-n} p_{i_n}$. This proves that $\phi_{\mathcal{P}}$ is computable over $X^{\mathcal{P}}$. Proposition 2.5.2 allows to conclude. $\square$

After the definition an important question is: are there computable partitions? the answer depends on the existence of open r.e sets with a zero-measure boundary.

Definition 3.1.3. A set A is said to be ***almost decidable*** if there are two r.e open sets U and V such that:

$$U \subset A, \quad V \subseteq A^{c}, \quad \mu(U) + \mu(V) = 1$$

remarks:

- a set is almost decidable if and only if its complement is almost decidable,
- an almost decidable set is always a continuity set,
- a μ-continuity ideal ball is always almost decidable,

- unless the space is disconnected (i.e. has non-trivial clopen subsets), no set can be *decidable*, i.e. semi-decidable (r.e) and with a semi-decidable complement (such a set must be clopen[1]). Instead, a set can be decidable *with probability* 1: there is an algorithm which decides if a point belongs to the set or not, for almost every point. That is why we call it *almost decidable*.

Ignoring computability, the existence of open μ-continuity sets directly follows from the fact that the collection of open sets is uncountable and μ is finite. The problem in the computable setting is that there are only countable many open r.e sets. Fortunately, there still always exists a basis of almost decidable balls. This result, first obtained in [HR07] with other techniques, will be used many times in the sequel, in particular it directly implies the existence of computable partitions. For completeness we present a different, self-contained proof.

Theorem 3.1.2. *There is a family of uniformly computable reals* $(r_n^i)_{i,n\in\mathbb{N}}$ *such that for all* i, $\{r_n^i : n \in \mathbb{N}\}$ *is dense in* $\mathbb{R}^+$ *and such that for every* i, n, *the ball* $B(s_i, r_n^i)$ *is almost decidable.*

Proof. Let s_i be an ideal point. Put $I_{\langle j,k\rangle} = [q_j, q_k]$ with q_j, q_k positive rational numbers. We show that for every $n = \langle j, k\rangle$ we can compute, uniformly in n, a real $r_n^i \in I_n$ for which $\mu(\partial B(s_i, r_n^i)) = 0$. First observe that for a closed interval $I = [a, b]$ $(a, b \in \mathbb{Q})$, the complement of $B_I = \overline{B}(s_i, b)/B(s_i, a)$, is r.e open. Then by corollary 2.5.2, its measure is lower semi-computable and then we can semi-decide for a given rational q the relation $\mu(B_I) < q$. The algorithm computing r_n^i enumerates a sequence of nested closed intervals $(J_k)_{k\in\mathbb{N}}$ whose length tends to 0, with $J_0 = I_n$, and such that for all k, $\mu(B_{J_k}) < 2^{-k+1}$. Then $\{r_n^i\} = \cap_{k\geq 1} J_k$. It works as follows:

In stage $k+1$ (the interval $J_k = [a, b]$ has already been found), put $m = \frac{b-a}{3}$ and test in parallel $\mu(B_{[a,a+m]}) < 2^{-k}$ and $\mu(B_{[b-m,b]}) < 2^{-k}$. Since $\mu(B_{J_k}) < 2^{-k+1}$, one of the tests must stop, and then provides the "good" interval J_{k+1} for which the condition holds. □

We denote by $B^{\langle i,n\rangle}$ the almost decidable ball $B(s_i, r_n^i)$.

The family $\{B^{\langle i,n\rangle} : i, n \in \mathbb{N}\}$ is a basis for the topology. It is even effectively equivalent to the basis of ideal balls : every ideal ball can be expressed as a r.e. union of almost decidable balls, and vice-versa.

We finish presenting some results that will be needed in the next subsection.

Corollary 3.1.1. *On every computable probability space, there exists a family of uniformly computable partitions which generates the Borel* σ*-algebra.*

Proof. Take $\mathcal{P}_{\langle i,n\rangle} = \{B(s_i, r_n^i), X \setminus \overline{B}(s_i, r_n^i)\}$ where $\overline{B}$ is the closed ball: as the almost decidable balls form a basis of the topology, the σ-algebra generated by the P_k is the Borel σ-field. □

Proposition 3.1.1. *If* A *is almost decidable then* $\mu(A)$ *is a computable real number.*

[1] In Cantor space for example (which is totally disconnected), every cylinder (ball) is a decidable set. Indeed, deciding if some infinite sequence belongs to some cylinder reduces to a finite pattern-matching.

Proof. Since U and V are r.e open, by theorem 2.5.2 their measures are lower-semi-computable. As $\mu(U)+\mu(V)=1$, their measures are also upper-semi-computable. □

The following regards the computability of inducing a measure in a subset and will be used in the proof of prop. 3.2.1

Proposition 3.1.2. *Let μ be a computable measure and A an almost decidable subset of X. Then the induced measure $\mu_A(.)=\mu(.|A)$ is computable. Furthermore, $R_{\mu_A}=R_\mu\cap A$.*

Proof. let $W=B_{n_1}\cup\ldots\cup B_{n_k}$ be a finite union of ideal balls. $\mu_A(W)=\mu(W\cap A)/\mu(A)=\mu(W\cap U)/\mu(A)$. $W\cap U$ is a r.e open set, so its measure is lower semi-computable. As $\mu(A)$ is computable, $\mu_A(W)$ is lower semi-computable. Note that everything is uniform in $\langle n_1,\ldots,n_k\rangle$. The result follows from theorem 2.5.2.

Let U and V as in the definition of an almost decidable set. First note that $R_\mu\cap A=R_\mu\cap U$, as $R_\mu\subseteq U\cup V$ by lemma 2.6.1. Again by lemma 2.6.1, $R_{\mu_A}\subseteq U$, and as $\mu_A\leq\frac{1}{\mu(A)}\mu$, every μ-effective null set is also a μ_A-effective null set, so $R_{\mu_A}\subseteq R_\mu$. Hence, we have $R_{\mu_A}\subseteq R_\mu\cap U$.

Conversely, $R^c_{\mu_A}$ being a μ_A-effective null set, its intersection with U is a μ-effective null set, by definition of μ_A. So $R^c_{\mu_A}\cap U\subseteq R^c_\mu$, which is equivalent to $R_\mu\cap U\subseteq R_{\mu_A}$. □

3.2. **Some statistical properties of random points.** With the tools developed so far, it is possible to translate many results of the form

$$\mu\{x:P(x)\}=1,$$

with P some predicate, into an "individual" result of the form:

"If x is μ-random, then $P(x)$".

In this section we give two examples: recurrence and statistical typicality.

Definition 3.2.1. Let X be a metric space. A point $x\in X$ is said to be ***recurrent*** for a transformation $T:X\to X$, if $\liminf_n d(x,T^nx)=0$.

Proposition 3.2.1 (Random points are recurrent)**.** *Let (X,μ) be a computable probability space. If x is μ-random, then it is recurrent with respect to* every *measure preserving endomorphism T on (X,μ).*

Proof. take $x\in R_\mu$ and B an almost decidable neighborhood of x. Then $\mu(B)>0$ and there is a r.e open set U such that:

$$\bigcup_{n\geq 1}T^{-n}B=U\cap D$$

where D is the domain of computability of T. By the Poincaré recurrence theorem, this set has full measure for $\mu_B(.)=\mu(.|B)$. By proposition 3.1.2, $x\in R_{\mu_B}$, so by lemma 2.6.1, x is in U. □

We now prove that random points satisfy a stronger property to be used in the sequel: statistical typicality. Let us then introduce this concept.

Let X be a metric space and T be a continuous transformation on X. Let $C_b(X)$ be the space of bounded real-valued continuous functions on X. For $f\in C_b(X)$ define:

$$\overline{f}(x) := \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) \tag{3}$$

at the points x where this limit exists. We recall that a point x is called ***generic*** for T if $\overline{f}(x)$ is defined for every $f \in C_b(X)$.

Every generic point x generates a probability measure μ_x which is invariant for T, dually defined by:

$$\int_X f d\mu_x = \overline{f}(x) \text{ for all } f \in C_b(X). \tag{4}$$

In other words, x is generic if the measure $\nu_n = \frac{1}{n}\sum_{j<n} \delta_{T^j x}$ converges weakly to μ_x, where δ_y is the Dirac probability measure concentrated on y. Let μ be an ergodic measure for T. A generic point x is said to be ***μ-typical*** if $\mu_x = \mu$. The well-known Birkhoff ergodic theorem says that for each ergodic measure μ, the set of μ-typical points has μ-measure one.

From a statistical point of view, μ-typical points are those whose orbits reproduce the main statistical features of μ (in particular they are a total measure set), hence in some sense they are *random* for the dynamic.

What algorithmically random points have to do with dynamically random points?

This problem has already been studied by V'yugin ([V'y97]) in the particular case of the Cantor space and for computable observables. We prove a general version which applies to computable dynamics on any computable probability space, for any bounded continuous (not necessarily computable) observable. The strategy is simple: we use computable partitions to construct effective symbolic models and use the following particular case of V'yugin's main theorem.

Lemma 3.2.1. *Let μ be a computable shift-invariant ergodic measure on the Cantor space $\{0,1\}^\omega$. Then for each μ-random sequence ω:*

$$\lim_n \frac{1}{n} \sum_{i=0}^{n} \omega_i = \mu([1]) \tag{5}$$

We are now able to prove:

Theorem 3.2.1. *Let (X,μ) be a computable probability space. Then each μ-random point x is μ-typical for* every *ergodic endomorphism T.*

We remark that the theorem holds uniformly for all bounded continuous observables and all ergodic endomorphisms.

Proof. Let f_A be the characteristic function of the set A. First, let us show that if A is an almost decidable set then for all μ-random point x:

$$\lim_n \frac{1}{n} \sum_{i=0}^{n} f_A \circ T^i(x) = \mu(A) \tag{6}$$

Indeed, consider the *computable partition* defined by $\mathcal{P} := \{U, V\}$ with U and V as in definition 3.1.3 and the associated symbolic model $(X_\mathcal{P}, \sigma, \mu_\mathcal{P})$. By proposition 3.1.1, $\phi_\mathcal{P}(x)$ is a well defined $\mu_\mathcal{P}$-random infinite sequence, so lemma 3.2.1 applies and gives (6). This can be reformulated as the convergence of $\nu_n(A)$ to $\mu(A)$. Now,

the collection of almost decidable sets satisfies proposition 2.5.1, so ν_n converges weakly to μ: x is μ-typical. □

4. Measure-theoretic entropies

Suppose discrete objects (symbolic strings for instance) are produced by some source. The tendency of the source toward producing such object more than such other can be modeled by a probability distribution, which gives more information than the crude set of possible outcomes. The Shannon entropy of the probabilistic source measures the degree of uncertainty that lasts when taking the probability distribution into account.

Any ergodic dynamical system (X, T, μ) can be seen as a source of outputs. Kolmogorov and Sinaï adapted Shannon's theory to dynamical systems in order to measure the degree of unpredictability or chaoticity of an ergodic system. The first step consists in discretizing the space X using finite partitions. Let $\xi = \{C_1, \ldots, C_n\}$ be a finite measurable partition of X. Then let $T^{-1}\xi$ be the partition whose atoms are the pre-images $T^{-1}C_i$. Then let

$$\xi_n = \xi \vee T^{-1}\xi \vee T^{-2}\xi \vee \ldots \vee T^{-(n-1)}\xi$$

be the partition given by the sets of the form

$$C_{i_0} \cap T^{-1}C_{i_1} \cap \ldots \cap T^{-(n-1)}C_{i_{n-1}},$$

varying C_{i_j} among all the atoms of ξ. Knowing which atom ξ_n a point x belongs to comes to knowing which atoms of the partition ξ the orbit of x visits up to time $n-1$.

The measure-theoretical entropy of the system w.r.t the partition ξ can then be thought as the rate (per time unit) of gained information (or removed uncertainty) when observations of the type "$T^n(x) \in C_i$" are performed. This is of great importance when classifying dynamical systems: it is a measure-theoretical invariant, which enables one to distinguish non-isomorphic systems.

We briefly recall the definition. For more details, we refer the reader to [Bil65], [Wal82], [Pet83], [HK95].

4.1. Entropy with Shannon information.

Given a partition ξ and a point x, $\xi(x)$ denotes the atom of the partition x belongs to. Let us consider the ***Shannon information function*** relative to the partition ξ_n (the information which is gained by observing that $x \in \xi_n(x)$),

$$I_\mu(x|\xi_n) := -\log \mu(\xi_n(x))$$

and its mean, the entropy of the partition ξ_n,

$$H_\mu(\xi_n) := \int_X I_\mu(.|\xi_n)d\mu = \sum_{C \in \xi_n} -\mu(C)\log\mu(C)$$

The ***measure-theoretical*** or ***Kolmogorov-Sinaï entropy*** of T relative to the partition ξ is defined as:

$$h_\mu(T, \xi) = \lim_{n\to\infty} \frac{1}{n} H_\mu(\xi_n).$$

(which exists and is an infimum, since the sequence $H_\mu(\xi_n)_n$ is sub-additive). With the Shannon information function, it is possible to define a kind of point-wise notion

of entropy with respect to a partition ξ:

$$\limsup_n \frac{1}{n} I_\mu(x|\xi_n).$$

This local entropy is related to the global entropy of the system by the celebrated Shannon-McMillan-Breiman theorem:

Theorem (Shannon-McMillan-Breiman)**.** *Let T be an ergodic endomorphism of the probability space $(X, \mathscr{B}, \mu)$ and ξ a finite measurable partition. Then for μ-almost every x,*

$$\lim_{n\to\infty} \frac{1}{n} I_\mu(x|\xi_n) = h_\mu(T, \xi). \tag{7}$$

The convergence also holds in $L^1(X, \mathscr{B}, \mu)$.

Now we suppress the partition-dependency: the ***Kolmogorov-Sinaï entropy*** of (X, T, μ) is

$$h_\mu(T) := \sup\{h_\mu(T, \xi) : \xi \text{ finite measurable partition}\}$$

We recall the following two results that we will need later. The first proposition follows directly from the definitions.

Proposition 4.1.1. *If $(\Sigma^{\mathbb{N}}, \mu_\xi, \sigma)$ is the symbolic model associated to (X, μ, T, ξ) then $h_\mu(T, \xi) = h_{\mu_\xi}(\sigma)$.*

The next proposition is taken from [Pet83]:

Proposition 4.1.2. *If $(\xi_i)_{i\in\mathbb{N}}$ is a family of finite measurable partitions which generates the Borel σ-field up to sets of measure 0, then $h_\mu(T) = \sup_i h_\mu(T, \xi_0 \vee \ldots \vee \xi_i)$.*

4.2. **Entropy with Kolmogorov information.** In this section, T is an endomorphism of the computable probability space (X, μ) and $\xi = \{C_1, \ldots, C_k\}$ is a computable partition. Let $(\Sigma^{\mathbb{N}}, \mu_\xi, \sigma)$ be the effective symbolic model of (X, μ, T, ξ) where $\Sigma = \{1, \ldots, k\}$ (see section 3.1).

Kolmogorov introduced his algorithmic information content (also called Kolmogorov complexity) as a quantity of information, on the same level as Shannon information. When the measure, the transformation and the partition are computable, it makes sense to define the algorithmic equivalents of the notions defined above. It turns out that the two points of view are strongly related.

An atom C of the partition ξ_n can then be seen as a word of length n on the alphabet Σ, which allows one to consider its Kolmogorov complexity $K(C)$. For those points whose all iterates are covered by ξ (they form a constructive dense G_δ of full measure), we define the ***Kolmogorov information function*** relative to the partition ξ_n:

$$\mathcal{I}(x|\xi_n) := K(\xi_n(x))$$

which is independent of μ. We can then define ***algorithmic entropy*** of the partition ξ_n as the mean of $\mathcal{I}$:

$$\mathcal{H}_\mu(\xi_n) := \int_X \mathcal{I}(.|\xi_n) d\mu = \sum_{C\in\xi_n} \mu(C)K(C).$$

We also define a local notion of algorithmic entropy, which we call symbolic orbit complexity:

Definition 4.2.1 (Symbolic orbit complexity)**.** Let T be an endomorphism of the computable probability space (X,μ). For any finite computable partition ξ, we define $\mathcal{K}_\mu(x,T|\xi) := \limsup_n \frac{1}{n}\mathcal{I}(x|\xi_n)$. Then, we can suppress the dependence on ξ by taking the supremum over all computable partitions:

$$\mathcal{K}_\mu(x,T) := \sup\{\mathcal{K}_\mu(x,T|\xi) : \xi \text{ computable partition}\}$$

As there are only countably many computable partitions, $\mathcal{K}_\mu(x,T)$ is defined almost everywhere (at least on random points).

The quantity $\mathcal{K}_\mu(x,T|\xi)$ was introduced by Brudno in [Bru83] without any computability restriction on the space, the measure nor the transformation. He proved:

Theorem 4.2.1 (Brudno)**.** *$\mathcal{K}_\mu(x,T|\xi) = h_\mu(T,\xi)$ for μ-almost every point.*

Remark 4.2.1. Already Brudno remarked that if x has not an eventually periodic orbit, by taking the supremum of $\mathcal{K}_\mu(x,T|\xi)$ over all – not necessarily computable – finite partitions ξ generally gives an infinite quantity, that is why Brudno did not go further (he did not have a computable structure at his disposal), and proposed a topological definition using open covers instead of partitions.

We will compare $\mathcal{K}_\mu$ and Brudno orbit complexity in section 4.5, we now show that the hypothesis of definition 4.2.1 enables one to derive Brudno's theorem in a rather simple manner.

The theory of randomness and Kolmogorov complexity on the space of symbolic sequences provides powerful results (theorem 2.7.3 and proposition 2.7.2) which enable to relate the algorithmic entropies $\mathcal{I}_\mu$ and $\mathcal{H}_\mu$ to the Shannon entropies I_μ and H_μ (inequalities (9), (11)). We recall these two results: if $\Sigma^{\mathbb{N}}$ is endowed with a computable probability measure ν, then for all $\omega \in \Sigma^{\mathbb{N}}$,

$$-\log \nu[\omega_{0..n-1}] - d_\nu(\omega) \quad \leq \quad K(\omega_{0..n-1}) \quad \overset{+}{<} \quad -\log \nu[\omega_{0..n-1}] + K(n) \tag{8}$$

where d_ν is the deficiency of randomness, which satisfies $\int_{\Sigma^{\mathbb{N}}} d_\nu d\nu < 1$ and is finite exactly on Martin-Löf random sequences (the constant in $\overset{+}{<}$ does not depend on ω and n, see section 2.7.1).

4.3. **Equivalence between local entropies.** Applying (8) to $\nu = \mu_\xi$ directly gives:

$$I_\mu(.|\xi_n) - d_\mu \circ \phi_\xi \quad \leq \quad \mathcal{I}(.|\xi_n) \quad \overset{+}{<} \quad I_\mu(.|\xi_n) + K(n) \tag{9}$$

where it is defined (almost everywhere, at least on random points). Every μ-Martin-Löf random point x is mapped by ϕ_ξ on a μ_ξ-Martin-Löf random sequence (see proposition 2.6.1), whose randomness deficiency is finite. It then follows that the local entropies using Shannon information and Kolmogorov information coincide on μ-random points:

Proposition 4.3.1.

$$\mathcal{K}_\mu(x,T|\xi) = \limsup_n \frac{1}{n} I(x|\xi_n) \quad \textit{for every } \mu\textit{-Martin-Löf random point } x \tag{10}$$

This equality together with the Shannon-McMillan-Breiman theorem (7) give directly Brudno's theorem (theorem 4.2.1).

Remark 4.3.1 (Equivalence between global entropies). Now, the Kolmogorov-Sinaï entropy, originally expressed using Shannon entropy, can be expressed using algorithmic entropy. Taking the mean in (9), one obtains:

$$H_\mu(\xi_n) - 1 \quad \leq \quad \mathcal{H}_\mu(\xi_n) \quad \overset{+}{<} \quad H_\mu(\xi_n) + K(n) \tag{11}$$

So,

$$h_\mu(T|\xi) = \lim_n \frac{H_\mu(\xi_n)}{n} = \lim_n \frac{\mathcal{H}_\mu(\xi_n)}{n}$$

As the collection of computable partitions is generating (see corollary 3.1.1), the Kolmogorov-Sinaï entropy of (X, μ, T) can be characterized by:

$$h_\mu(T) = \sup \left\{ \lim_n \frac{\mathcal{H}_\mu(\xi_n)}{n} : \xi \text{ finite computable partition} \right\}.$$

It then follows that $\mathcal{K}_\mu(x, T) = h_\mu(T)$ for μ-almost every x. We now strengthen this, proving that it holds for all Martin-Löf random points.

4.4. **Orbit complexity vs entropy.** On the Cantor space, V'yugin ([V'y97]) and later Nakamura ([Nak05]) proved a slightly weaker version of the Shannon-McMillan-Breiman for Martin-Löf random sequences. In particular, we will use:

Theorem 4.4.1 (V'yugin)**.** *Let μ be a computable shift-invariant ergodic measure on $\Sigma^{\mathbb{N}}$. Then, for any μ-Martin-Löf random sequence ω,*

$$\limsup_{n\to\infty} -\frac{1}{n} \log \mu([\omega_{0..n-1}]) = h_\mu(\sigma).$$

Note that it is not known yet if the limit exists for all random sequences.

Using effective symbolic models, this can be easily extended to any computable probability space.

Corollary 4.4.1 (Shannon-McMillan-Breiman for random points)**.** *Let T be an ergodic endomorphism of the computable probability space (X, μ), and ξ a computable partition. For every μ-Martin-Löf random point x,*

$$\limsup_{n\to\infty} -\frac{1}{n} \log \mu(\xi_n(x)) = h_\mu(T, \xi).$$

Proof. Since ξ is computable, the symbolic model (X_ξ, μ_ξ, σ) is effective. Every μ-Martin-Löf random point x is mapped to a μ_ξ-Martin-Löf random sequence ω, for which the preceding theorem holds. Using the facts that $\mu(\xi_n(x)) = \mu_\xi([\omega_{0..n-1}])$ and $h_\mu(T, \xi) = h_{\mu_\xi}(\sigma)$ allows to conclude. □

Finally, this implies our first announced result:

Theorem 4.4.2. *Let T be an ergodic endomorphism of the computable probability space (X, μ). For every μ-Martin-Löf random point x:*

$$\mathcal{K}_\mu(x, T) = h_\mu(T).$$

Proof. We combine equality (10) and corollary 4.4.1: for every random point x, $\mathcal{K}_\mu(x, T|\xi) = \limsup_n \frac{1}{n} I_\mu(x|\xi_n) = h_\mu(T, \xi)$. Since the collection of all computable partitions generates the Borel σ-field (corollary 3.1.1), $\mathcal{K}_\mu(x, T) = \sup\{h_\mu(T, \xi) : \xi \text{ computable partition}\} = h_\mu(T)$ (proposition 4.1.2). □

4.5. **Orbit complexity.** In this section, $(X, d, \mathcal{S})$ is a computable metric space and $T : X \to X$ a transformation (for the moment, no continuity or computability assumption is put on T). We will consider a notion of orbit complexity which quantifies the algorithmic information needed to describe the orbit of x with finite but arbitrarily accurate precision. The definition we will give coincide with Brudno's original definition on compact spaces (see [Gal00]).

Given $\epsilon > 0$ and $n \in \mathbb{N}$, the algorithmic information that is needed to list a sequence of ideal points which follows the orbit of x for n steps at a distance less than ϵ is:

$$\mathcal{K}_n(x, T, \epsilon) := \min\{K(i_0, \ldots, i_{n-1}) : d(s_{i_j}, T^j x) < \epsilon \text{ for } j = 0, \ldots, n-1\}$$

where K is the self-delimiting Kolmogorov complexity.

We then define the maximal and minimal growth-rates of this quantity:

$$\begin{aligned} \overline{\mathcal{K}}(x, T, \epsilon) &:= \limsup_{n\to\infty} \frac{1}{n} \mathcal{K}_n(x, T, \epsilon) \\ \underline{\mathcal{K}}(x, T, \epsilon) &:= \liminf_{n\to\infty} \frac{1}{n} \mathcal{K}_n(x, T, \epsilon). \end{aligned}$$

As ϵ tends to 0, these quantities increase (or at least do not decrease), hence they have limits (which can be infinite).

Definition 4.5.1. The ***upper*** and ***lower orbit complexities*** of x under T are defined by:

$$\begin{aligned} \overline{\mathcal{K}}(x, T) &:= \lim_{\epsilon\to 0^+} \overline{\mathcal{K}}(x, T, \epsilon) \\ \underline{\mathcal{K}}(x, T) &:= \lim_{\epsilon\to 0^+} \underline{\mathcal{K}}(x, T, \epsilon). \end{aligned}$$

Remark 4.5.1. If T is computable, and assuming that ϵ takes only rational values, the n first iterates of x could be ϵ-shadowed by the orbit of a single ideal point instead of a pseudo-orbit of n ideal points. Actually it is easy to see that it gives the same quantities $\overline{\mathcal{K}}(x, T, \epsilon)$ and $\underline{\mathcal{K}}(x, T, \epsilon)$: let $\mathcal{K}'_n(x, T, \epsilon) = \min\{K(i) : d(T^j s_i, T^j x) < \epsilon \text{ for } j < n\}$, one has:

$$\begin{aligned} \mathcal{K}'_n(x, T, 2\epsilon) &\overset{+}{<} \mathcal{K}_n(x, T, \epsilon) + K(\epsilon) \\ \mathcal{K}_n(x, T, \epsilon) &\overset{+}{<} \mathcal{K}'_n(x, T, \epsilon/2) + K(n, \epsilon) \end{aligned}$$

Indeed, from ϵ and $i_0, \ldots, i_{n-1}$ some ideal point can be algorithmically found in the constructive open set $B(s_{i_0}, \epsilon) \cap \ldots \cap T^{-(n-1)} B(s_{i_{n-1}}, \epsilon)$, uniformly in $i_0, \ldots, i_{n-1}$. Its n first iterates 2ϵ-shadow the orbit of x, which proves the first inequality. For the second inequality, some $i_0, \ldots, i_{n-1}$ can be algorithmically found from n, ϵ, and a point s_i whose n first iterates $\epsilon/2$-shadow the orbit of x, taking any $s_{i_j} \in B(T^j s_i, \epsilon/2)$.

Remark 4.5.2. Under the same assumptions, one could define $K(B_n(s_i, \epsilon))$ to be $K(i, n, \epsilon)$, and replace $K(i)$ by $K(B_n(s_i, \epsilon))$ in the definition of $\mathcal{K}'_n(x, T, \epsilon)$, without changing the quantities $\overline{\mathcal{K}}(x, T, \epsilon)$ and $\underline{\mathcal{K}}(x, T, \epsilon)$. Indeed,

$$K(i) \overset{+}{<} K(B_n(s_i, \epsilon)) \overset{+}{<} K(i) + K(n) + K(\epsilon)$$

5. Equivalence of the two notions of orbit complexity for random points

We now prove:

Theorem 5.0.1. *Let T be an ergodic endomorphism of the computable probability space (X, μ), where X is compact. Then for every Martin-Löf random point x,*

$$\overline{\mathcal{K}}(x, T) = \mathcal{K}_\mu(x, T).$$

Proof of $\overline{\mathcal{K}}(x,T) \leq \mathcal{K}_\mu(x,T)$. Let $\epsilon > 0$. Choose a computable partition ξ of diameter $< \epsilon$ (this is why we require X to be compact). To every cell of ξ, associate an ideal point which is inside (as ξ is computable, this can be done in a computable way, but we actually do not need that). The translation of symbolic sequences in sequences of ideal points through this finite dictionary is constructive, and transforms the symbolic orbit of a point x into a sequence of ideal points which is ϵ-close to the orbit of x. So $\overline{\mathcal{K}}(x, T, \epsilon) \leq \mathcal{K}_\mu(x, T|\xi)$. The inequality follows letting ϵ tend to 0. □

To prove the other inequality, we recall some technical stuff. The self-delimiting Kolmogorov complexity of natural numbers $k \geq 1$ satisfies

$$K(k) \overset{+}{<} f(k)$$

where $f(x) = \log x + 1 + 2\log(\log x + 1)$ for all $x \in \mathbb{R}, x \geq 1$. f is a concave increasing function and $x \mapsto xf(1/x)$ is an increasing function on $]0, 1]$ which tends to 0 as $x \to 0$.

We recall that for finite sequences of natural numbers $(k_1, \ldots, k_n)$, one has

$$K(k_1, \ldots, k_n) \overset{+}{<} K(k_1) + \ldots + K(k_n)$$

as the shortest descriptions for $k_1, \ldots, k_n$ can be extracted from their concatenation (this is one reason to use the self-delimiting complexity instead of the plain complexity).

Lemma 5.0.1. *Let Σ be a finite alphabet and $n \in \mathbb{N}$. Let $u, v \in \Sigma^n$ and $0 < \alpha < 1/2$ such that the density of the set of positions where u and v differ is less than α, that is:*

$$\frac{1}{n}\#\{i \leq n : u_i \neq v_i\} < \alpha < 1/2$$

Then $\left|\frac{1}{n}K(u) - \frac{1}{n}K(v)\right| < \alpha f(1/\alpha) + \frac{c}{n}$ where c is a constant independent of u, v and n.

Proof. Let $(i_1, \ldots, i_p)$ be the ordered sequence of indices where u and v differ. By hypothesis, $p/n < \alpha$. Put $k_1 = i_1$ and $k_j = i_j - i_{j-1}$ for $2 \leq j \leq p$.

We now show that u can be recovered from v and roughly $\alpha f(1/\alpha)n$ bits more. Indeed u can be computed from $(v, k_1, \ldots, k_p)$, constructing the string which coincides with v everywhere but at positions $k_1, k_1 + k_2, \ldots, k_1 + \ldots + k_p$, so $K(u) \overset{+}{<} K(v) + K(k_1) + \ldots + K(k_p) \overset{+}{<} K(v) + f(k_1) + \ldots + f(k_p)$.

Now, as f is a concave increasing function, one has:

$$\frac{1}{p}\sum_{j \leq p} f(k_j) \leq f\left(\frac{1}{p}\sum_{j \leq p} k_j\right) = f\left(\frac{i_p}{p}\right) \leq f\left(\frac{n}{p}\right)$$

As a result,

$$\frac{1}{n}K(u) \leq \frac{1}{n}K(v) + \frac{p}{n}f\left(\frac{n}{p}\right) + \frac{c}{n}$$

where c is some constant independent of u, v, n, p. As $p/n < \alpha < 1/2$ and $x \mapsto xf(1/x)$ is increasing for $x \leq 1/2$, one has:

$$\frac{1}{n}K(u) \leq \frac{1}{n}K(v) + \alpha f(1/\alpha) + \frac{c}{n}$$

Switching u and v gives the result (c may be changed). □

We are now able to prove the other inequality.

Proof of $\mathcal{K}_\mu(x,T) \leq \overline{\mathcal{K}}(x,T)$. Fix some computable partition ξ. We show that for any $\beta > 0$ there is some $\epsilon > 0$ such that for every Martin-Löf random point x, $\mathcal{K}_\mu(x,T|\xi) \leq \overline{\mathcal{K}}(x,T,\epsilon) + \beta$. As $\overline{\mathcal{K}}(x,T,\epsilon)$ increases as $\epsilon \to 0^+$ and β is arbitrary, the inequality follows.

First take $\alpha < 1/2$ such that $\alpha f(1/\alpha) < \beta$, and remark that

$$\lim_{\epsilon \to 0^+} \mu\left(\overline{(\partial\xi)^\epsilon}\right) = \mu(\partial\xi) = 0$$

Hence there is some ϵ such that $\mu\left(\overline{(\partial\xi)^{2\epsilon}}\right) < \alpha$. From a sequence of ideal points we will reconstruct the symbolic orbit of a random point with a density of errors less than α. Lemma 5.0.1 will then allow to conclude.

We define an algorithm $\mathcal{A}(\epsilon, i_0, \dots, i_{n-1})$ with $\epsilon \in \mathbb{Q}_{>0}$ and $i_0, \dots, i_{n-1} \in \mathbb{N}$ which outputs a word $a_0 \dots a_{n-1}$ on the alphabet ξ. To compute a_j, $\mathcal{A}$ semi-decides in a dovetail picture:

- $s_{i_j} \in C$ for every $C \in \xi$,
- $s \in C$ for every $s \in B(s_{i_j}, \epsilon)$ and every $C \in \xi$.

The first test which stops provides some $C \in \xi$: put $a_j = C$.

Let x be a random point whose iterates are covered by ξ, and $s_{i_0}, \dots, s_{i_{n-1}}$ be ideal points which ϵ-shadow the first n iterates of x. We claim that $\mathcal{A}$ will halt on $(\epsilon, i_0, \dots, i_{n-1})$. Indeed, as $T^j x$ belongs to some $C \in \xi$, $C \cap B(s_{i_j}, \epsilon)$ is a non-empty open set and then contains at least one ideal point s, which will be eventually dealt with.

We now compare the symbolic orbit of x with the symbolic sequence computed by $\mathcal{A}$. A discrepancy at rank j can appear only if $T^j x \in (\partial\xi)^{2\epsilon}$. Indeed, if $T^j x \notin (\partial\xi)^{2\epsilon}$ then $B(T^j x, 2\epsilon) \subseteq C$ where C is the cell $T^j x$ belongs to. As $d(s_{i_j}, T^j x) < \epsilon$, $B(s_{i_j}, \epsilon) \subseteq B(x, 2\epsilon) \subseteq C$, so the algorithm gives the right cell.

Now, as x is typical,

$$\limsup_{n \to \infty} \frac{1}{n} \#\{j < n : T^j x \in (\partial\xi)^{2\epsilon}\} \leq \mu\left(\overline{(\partial\xi)^{2\epsilon}}\right) < \alpha$$

so there is some n_0 such that for all $n \geq n_0$, $\frac{1}{n}\#\{j < n : T^j x \in (\partial\xi)^{2\epsilon}\} < \alpha$. This implies that for all $n \geq n_0$ and ideal points $s_{i_0}, \dots, s_{i_{n-1}}$ which ϵ-shadow the first n iterates of x and with minimal complexity, the algorithm $\mathcal{A}(\epsilon, i_0, \dots, i_{n-1})$ produces a symbolic string u which differs from the symbolic orbit v of x of length n with a density of errors $< \alpha$. As $K(u) \stackrel{+}{<} K(\epsilon) + \mathcal{K}_n(x, T, \epsilon)$ and $\alpha f(1/\alpha) < \beta$,

applying lemma 5.0.1 gives:

$$\begin{aligned}\frac{1}{n}K(\xi_n(x)) = \frac{1}{n}K(v) &\leq \frac{1}{n}K(u) + \alpha f(1/\alpha) + \frac{c}{n} \\ &\leq \frac{1}{n}\left(\mathcal{K}_n(x,T,\epsilon) + K(\epsilon) + c'\right) + \beta + \frac{c}{n}\end{aligned}$$

where c' is independent of n. Taking the $\limsup$ as $n \to \infty$ gives:

$$\mathcal{K}_\mu(x,T|\xi) \leq \overline{\mathcal{K}}(x,T,\epsilon) + \beta$$

□

Combining theorems 5.0.1 and 4.4.2, we obtain a version of Brudno's theorem (theorem 1) for Martin-Löf random points.

Corollary 5.0.1. *Let T be an ergodic endomorphism of the computable probability space (X,μ), where X is compact. Then for every Martin-Löf random point x:*

$$\overline{\mathcal{K}}(x,T) = h_\mu(T)$$

6. Topological entropies

Bowen's definition of topological entropy is reminiscent of the capacity (or box counting dimension) of a totally bounded subset of a metric space. In order to find relations with orbit complexity we will also use another characterization of topological entropy, expressing it as a kind of Hausdorff dimension. We first present Bowen's definition.

In this section, X is a metric space and $T : X \to X$ a continuous map.

6.1. Entropy as a capacity.

We recall the definition: for $n \geq 0$, let us define the distance $d_n(x,y) = \max\{d(T^ix,T^iy) : 0 \leq i < n\}$ and the Bowen ball $B_n(x,\epsilon) = \{y : d_n(x,y) < \epsilon\}$, which is open by continuity of T. Given a totally bounded set $Y \subseteq X$ and numbers $n \geq 0, \epsilon > 0$, let $N(Y,n,\epsilon)$ be the minimal cardinality of a cover of Y by Bowen balls $B_n(x,\epsilon)$. A set of points E such that $\{B_n(x,\epsilon) : x \in E\}$ is a cover of Y is also called an (n,ϵ)-spanning set of Y. One then defines:

$$h_1(T,Y,\epsilon) = \limsup_{n\to\infty} \frac{\log N(Y,n,\epsilon)}{n}$$

which is non-decreasing as $\epsilon \to 0$, so the following limit exists:

$$h_1(T,Y) = \lim_{\epsilon\to 0} h_1(Y,T,\epsilon).$$

When X is compact, the ***topological entropy*** of T is $h(T) = h_1(T,X)$. It measures the exponential growth-rate of the number of distinguishable orbits of the system.

Remark 6.1.1. The topological entropy can be defined using separated sets instead of open covers: a subset A of X is (n,ϵ)-separated if for any distinct points $x,y \in A$, $d_n(x,y) > \epsilon$. Let us define $M(Y,n,\epsilon)$ as the maximal cardinality of an (n,ϵ)-separated subset of Y. It is easy to see that $M(Y,n,2\epsilon) \leq N(Y,n,\epsilon) \leq M(Y,n,\epsilon)$, and hence $h_1(T,Y)$ can be alternatively defined using $M(Y,n,\epsilon)$ in place of $N(Y,n,\epsilon)$.

6.2. **Entropy as a dimension.** It is possible to define a topological entropy which is an analog of Hausdorff dimension. His definition coincides with the classical one in the compact case. Hausdorff dimension has stronger stability properties than box dimension, which has important consequences, as we will see in what follows. We refer the reader to [Pes98], [HK02] for more details.

Let X be a metric space and $T : X \to X$ a continuous map. The ϵ-size of $E \subseteq X$ is 2^{-s} where

$$s = \sup\{n \geq 0 : \mathrm{diam}(T^i E) \leq \epsilon \text{ for } 0 \leq i < n\}.$$

It measures how long the orbits starting from E are ϵ-close. As ϵ decreases, the ϵ-size of E is non-decreasing. The 2ϵ-size of a Bowen ball $B_n(x, \epsilon)$ is less than 2^{-n}.

In a way that is reminiscent from the definition of Hausdorff measure, let us define

$$m^s_\delta(Y, \epsilon) = \inf_{\mathcal{G}} \left\{ \sum_{U \in \mathcal{G}} (\epsilon\text{-size}(U))^s \right\}$$

where the infimum is taken over all countable covers $\mathcal{G}$ of Y by open sets of ϵ-size $< \delta$. This quantity is monotonically increasing as δ tends to 0, so the limit $m^s(Y, \epsilon) := \lim_{\delta \to 0^+} m^s_\delta(Y, \epsilon)$ exists and is a supremum. There is a critical value s_0 such that $m^s(Y, \epsilon) = \infty$ for $s < s_0$ and $m^s(Y, \epsilon) = 0$ for $s > s_0$. Let us define $h_2(T, Y, \epsilon)$ as this critical value:

$$h_2(T, Y, \epsilon) := \inf \{s : m^s(Y, \epsilon) = 0\} = \sup \{s : m^s(Y, \epsilon) = \infty\}.$$

As less and less covers are allowed when $\epsilon \to 0$ (the ϵ-size of sets does not decrease), the following limit exists

$$h_2(T, Y) := \lim_{\epsilon \to 0^+} h_2(T, Y, \epsilon)$$

and is a supremum. In [Pes98], it is proved that:

Theorem 6.2.1. *When Y is a T-invariant compact set, $h_1(T, Y) = h_2(T, Y)$.*

In particular, if the space X is compact, then $h(T) = h_1(T, X) = h_2(T, X)$.

6.3. **Orbit complexity vs entropy.** Now we prove the main theorem of the section:

Theorem 6.3.1 (Topological entropy vs orbit complexity)**.** *Let X be a compact computable metric space, and $T : X \to X$ a computable map. Then*

$$h(T) = \sup_{x \in X} \underline{\mathcal{K}}(x, T) = \sup_{x \in X} \overline{\mathcal{K}}(x, T).$$

In order to prove this theorem, we define an effective version of the topological entropy, which is strongly related to the complexity of orbits.

6.3.1. *Effective entropy as an effective dimension.* Before defining an effective version, we give a simple characterization which will accommodate to effectivisation.

Definition 6.3.1. A ***null s-cover*** of $Y \subseteq X$ is a set $E \subseteq \mathbb{N}^3$ such that:

(1) $\sum_{(i,n,p) \in E} 2^{-sn} < \infty$,
(2) for each $k, p \in \mathbb{N}$, the set $\{B_n(s_i, 2^{-p}) : (i, n, p) \in E, n \geq k\}$ is a cover of Y.

The idea is simple: every null s-cover induces open covers of arbitrary small size and arbitrary small weight. Remark that any null s-cover of Y is also a null s'-cover for all $s' > s$.

Lemma 6.3.1. *$h_2(T,Y)=\inf\{s : Y$ has a null s-cover$\}$.*

Proof. Suppose $s > h_2(T,Y)$. We fix $p,k\in\mathbb{N}$ and put $\epsilon = 2^{-p}$ and $\delta = 2^{-k}$. As $m^s_\delta(Y,\epsilon)=0$, there is a cover $(U_{j,k,p})_j$ of Y by open sets of ϵ-size $\delta_{j,k,p}<\delta$ with $\sum_j \delta^s_{j,k,p} < 2^{-(k+p)}$. Let s_i be any ideal point in $U_{j,k,p}$. If $\delta_{j,k,p}>0$, then $\delta_{j,k,p}=2^{-n}$ for some n. If $\delta_{j,k,p}=0$, take any $n\geq (j+k+p)/s$. In both cases, $U_{j,k,p}$ is included in the Bowen ball $B_n(s_i,\epsilon)$. We define $E_{k,p}$ as the set of (i,n,p) obtained this way, and $E=\bigcup_{k,p}E_{k,p}$. By construction, for each k,p, $\{B_n(s_i,2^{-p}) : (i,n,p)\in E, n\geq k\}$ is a cover of Y. Moreover, $\sum_{(i,n,p)\in E_{k,p}} 2^{-sn}\leq \sum_j \delta^s_{j,k,p}+\sum_j 2^{-(j+k+p)}\leq 2^{-(k+p)+2}$, so $\sum_{(i,n,p)\in E}2^{-sn}<\infty$.

Conversely, if Y has a null s-cover E, take $\epsilon,\delta>0$ and p,k such that $\epsilon > 2^{-p+1}$ and $\delta>2^{-k}$. For all $k'\geq k$, the family $\{B_n(s_i,2^{-p}) : (i,n,p)\in E, n\geq k'\}$ is a cover of Y by open sets of ϵ-size smaller than $2^{-n}\leq\delta$. Moreover, $\sum_{(i,n,p)\in E, n\geq k'}2^{-sn}$ tends to 0 as k' grows, so $m^s_\delta(Y,\epsilon)=0$. It follows that $s\geq h_2(T,Y)$. □

By an *effective* null s-cover, we mean a null s-cover E which is a r.e. subset of $\mathbb{N}^3$.

Definition 6.3.2. The ***effective topological entropy*** of T on Y is defined by

$$h_2^{\text{eff}}(T,Y)=\inf\{s : Y \text{ has an effective null } s\text{-cover}\}$$

As less null s-covers are allowed in the effective version, $h_2(T,Y)\leq h_2^{\text{eff}}(T,Y)$. Of course, if $Y\subseteq Y'$ then $h_2^{\text{eff}}(T,Y)\leq h_2^{\text{eff}}(T,Y')$. We now prove:

Theorem 6.3.2 (Effective topological entropy vs lower orbit complexity)**.** *Let X be an effective metric space and $T: X\to X$ a continuous map. For all $Y\subseteq X$,*

$$h_2^{\text{eff}}(T,Y)=\sup_{x\in Y}\underline{\mathcal{K}}(x,T)$$

which implies in particular that $h_2^{\text{eff}}(T,\{x\})=\underline{\mathcal{K}}(x,T)$: the restriction of the system to a single orbit may have *positive* effective topological entropy.

This kind of result has already been obtained for the Hausdorff dimension of subsets of the Cantor space, proving that the effective dimension of a set A is the supremum of the lower growth-rate of Kolmogorov complexity of sequences in A (which corresponds to theorem 6.3.2 for sub-shifts). This remarkable property is a counterpart of the countable stability property of Hausdorff dimension ($\dim Y = \sup_i \dim Y_i$ when $\bigcup_i Y_i = Y$) (see [CH94], [May01], [Lut03], [Rei04], [Sta05]).

Theorem 6.3.2 is a direct consequence of the two following lemmas.

Lemma 6.3.2. *Let $\alpha\geq 0$ and $Y_\alpha=\{x:\underline{\mathcal{K}}(x,T)\leq\alpha\}$. One has $h_2^{\text{eff}}(T,Y_\alpha)\leq\alpha$.*

Proof. Let $\beta>\alpha$ be a rational number. We define the r.e. set $E=\{(i,n,p) : K(i,n,p)<\beta n\}$. Let $p\in\mathbb{N}$ and $\epsilon=2^{-p}$. If $x\in Y_\alpha$ then $\underline{\mathcal{K}}(x,T,\epsilon)\leq\alpha<\beta$ so for infinitely many n, there is some s_i such that $x\in B_n(s_i,\epsilon)$ and $K(i,n,p)<\beta n$. So for all k, $\{B_n(s_i,2^{-p}) : (i,n,p)\in E, n\geq k\}$ covers Y_α. Moreover, $\sum_{(i,n,p)\in E}2^{-\beta n}\leq\sum_{(i,n,p)\in E}2^{-K(i,n,p)}\leq 1$.

E is then an effective null β-cover of Y_α, so $h_2^{\text{eff}}(T,Y_\alpha)\leq\beta$. And this is true for every rational $\beta>\alpha$. □

Lemma 6.3.3. *Let $Y\subseteq X$. For all $x\in Y$, $\underline{\mathcal{K}}(x,T)\leq h_2^{\text{eff}}(T,Y)$.*

Proof. Let $s > h_2^{\mathrm{eff}}(T,Y)$: Y has an effective null s-cover E. As $\sum_{(i,n,p)\in E} 2^{-sn} < \infty$, by the coding theorem $K(i,n,p) \leq sn + c$ for some constant x, which does not depend on i,n,p. If $x \in Y$, then for each p,k, x is in a ball $B_n(s_i, 2^{-p})$ for some $n \geq k$ with $(i,n,p) \in E$. Then $\mathcal{K}_n(x,T,2^{-p}) \leq sn + c$ for infinitely many n, so $\underline{\mathcal{K}}(x,T,2^{-p}) \leq s$. As this is true for all p, $\underline{\mathcal{K}}(x,T) \leq s$. As this is true for all $s > h_2^{\mathrm{eff}}(T,Y)$, we can conclude. □

Proof of theorem 6.3.2. By lemma 6.3.3, $\alpha := \sup_{x\in Y} \underline{\mathcal{K}}(x,T) \leq h_2^{\mathrm{eff}}(T,Y)$. Now, as $Y \subseteq Y_\alpha$, $h_2^{\mathrm{eff}}(T,Y) \leq h_2^{\mathrm{eff}}(T,Y_\alpha) \leq \alpha$ by lemma 6.3.2. □

The definition of an effective null α-cover involves a summable computable sequence. The universality of the sequence $2^{-K(i)}$ among summable lower semi-computable sequences is at the core of the proof of the preceding theorem, which states that there is a universal effective null α-cover, for every $\alpha \geq 0$. In other words, there is a maximal set of effective topological entropy $\leq \alpha$, and this set is $Y_\alpha = \{x \in X : \underline{\mathcal{K}}(x,T) \leq \alpha\}$.

The definition of the topological entropy as a capacity could be also made effective, restricting to effective covers. Classical capacity does not share with Hausdorff dimension the countable stability. For the same reason, its effective version is not related with the orbit complexity as strongly as the effective topological entropy is. Nevertheless, a weaker relation holds, which is sufficient for our purpose: the upper complexity of orbits is bounded by the effective capacity. We do not develop this and only state the needed property (which implicitly uses the fact that the effective capacity coincides with the classical capacity for a compact computable metric space):

Lemma 6.3.4. *Let X be a compact computable metric space. For all $x \in X$, $\overline{\mathcal{K}}(x,T) \leq h_1(T,X)$.*

Proof. We first construct a r.e. set $E \subseteq \mathbb{N}^3$ such that for each n, p, $\{s_i : (i,n,p) \in E\}$ is a $(n, 2^{-p})$-spanning set and a $(n, 2^{-p-2})$-separated set. Let us fix n and p and enumerate $E_{n,p} = \{i : (i,n,p) \in E\}$, in a uniform way. The algorithm starts with $S = \emptyset$ and $i = 0$. At step i it analyzes s_i and decides to add it to S or not, and goes to step $i+1$. $E_{n,p}$ is the set of points which are eventually added to S.

> **Step *i*:** for each ideal point $s \in S$, test in parallel $d_n(s_i, s) < 2^{-p-1}$ and $d_n(s_i, s) > 2^{-p-2}$: at least one of them must stop. If the first one stops first, reject s_i and go to Step $i+1$. If the second one stops first, go on with the other points $s \in S$: if all S has been considered, then add s_i to S and go to Step $i+1$.

By construction, the set of selected ideal points forms a $(n, 2^{-p-2})$-separated set. If there is $x \in X$ which is at distance at least 2^{-p} from every selected point, then let s_i be an ideal point s_i with $d_n(x, s_i) < 2^{-p-1}$: s_i is at distance at least 2^{-p-1} from every selected point, so at step i it must have been selected, as the first test could not stop. This is a contradiction: the selected points form a $(n, 2^{-p})$-spanning set.

From the properties of $E_{n,p}$ it follows that $N(X, n, 2^{-p}) \leq |E_{n,p}| \leq M(X, n, 2^{-p-2})$, and then

$$\sup_p \left(\limsup \frac{1}{n} \log |E_{n,p}| \right) = h_1(T,X)$$

If $\beta > h_1(T,X)$ is a rational number, then for each p, there is $k \in \mathbb{N}$ such that $\log |E_{n,p}| < \beta n$ for all $n \geq k$.

Now, for $s_i \in E_{n,p}$, $K(i) \stackrel{+}{<} \log|E_{n,p}| + 2\log\log|E_{n,p}| + K(n,p)$ by proposition 2.7.1. Take $x \in X$: x is in some $B_n(s_i, 2^{-p})$ for each n, so $\overline{\mathcal{K}}(x,T,2^{-p}) \leq \limsup_n \frac{1}{n}\log|E_{n,p}| \leq \beta$ as $\log|E_{n,p}| < \beta n$ for all $n \geq k$. As this is true for all p and all $\beta > h_1(T,X)$, $\overline{\mathcal{K}}(x,T) \leq h_1(T,X)$ and this for all $x \in X$. □

We are now able to prove theorem 6.3.1. Combining the several results established above:

$$\underset{\text{(theorem 6.2.1)}}{h_1(T,X) = h_2(T,X)} \quad \leq \quad \underset{\text{(theorem 6.3.2)}}{h_2^{\text{eff}}(T,X) = \sup_{x\in X} \underline{\mathcal{K}}(x,T)} \quad \leq \quad \underset{\text{(lemma 6.3.4)}}{\sup_{x\in X} \overline{\mathcal{K}}(x,T) \leq h_1(T,X)}$$

and the statement is proved.

References

[Bil65] Patrick Billingsley. *Ergodic Theory and Information*. John Wiley, New York, 1965.

[Bil68] Patrick Billingsley. *Convergence of Probability Measures*. John Wiley, New York, 1968.

[Bru83] A.A. Brudno. Entropy and the complexity of the trajectories of a dynamical system. *Transactions of the Moscow Mathematical Society*, 2:127–151, 1983.

[CH94] Jin-Yi Cai and Juris Hartmanis. On Hausdorff and topological dimensions of the Kolmogorov complexity of the real line. In *Proceedings of the 30th IEEE symposium on Foundations of computer science*, pages 605–619, Orlando, FL, USA, 1994. Academic Press, Inc.

[Gác05] Peter Gács. Uniform test of algorithmic randomness over a general space. *Theoretical Computer Science*, 341:91–137, 2005.

[Gal00] Stefano Galatolo. Orbit complexity by computable structures. *Nonlinearity*, 13:1531–1546, 2000.

[GHR07a] Stefano Galatolo, Mathieu Hoyrup, and Cristóbal Rojas. Computable invariant measures. *In preparation*, 2007.

[GHR07b] Stefano Galatolo, Mathieu Hoyrup, and Cristóbal Rojas. An effective Borel-Cantelli lemma. Constructing orbits with required statistical properties. *Submitted to JEMS*, 2007. Available on arxiv.

[HK95] Boris Hasselblatt and Anatole Katok. *Introduction to the Modern Theory of Dynamical Systems*, volume 54 of *Encyclopedia of Mathematics and Its Applications*. Cambridge University Press, 1995.

[HK02] Boris Hasselblatt and Anatole Katok. *Handbook of Dynamical Systems*, volume 1A, chapter Principle Structures. Elsevier, 2002.

[HR07] Mathieu Hoyrup and Cristóbal Rojas. Computability of probability measures and Martin-Löf randomness over metric spaces. *Submitted to Information and Computation*, 2007. Available on arxiv.

[Lut03] Jack H. Lutz. The dimensions of individual strings and sequences. *Information and Computation*, 187(1):49–79, 2003.

[LV93] Ming Li and Paul M. B. Vitanyi. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer-Verlag, Berlin, 1993.

[May01] Elvira Mayordomo. A Kolmogorov complexity characterization of constructive Hausdorff dimension. *Electronic Colloquium on Computational Complexity (ECCC)*, 8(059), 2001.

[Nak05] Masahiro Nakamura. Ergodic theorems for algorithmically random sequences. *Proceedings of the Symposium on Information Theory and Its Applications*, 2005.

[Pes98] Yakov B. Pesin. *Dimension Theory in Dynamical Systems: Contemporary Views and Applications*. Chicago Lectures in Mathematics. The University of Chicago Press, 1998.

[Pet83] Karl Petersen. *Ergodic Theory*. Cambridge Univ. Press, 1983.

[Rei04] Jan Reimann. *Computability and fractal dimension*. PhD thesis, Ruprecht-Karls Universität Heidelberg, 2004.

[Rog87] Hartley Jr. Rogers. *Theory of Recursive Functions and Effective Computability*. MIT Press, Cambridge, MA, USA, 1987.

[Sta05] Ludwig Staiger. Constructive dimension equals Kolmogorov complexity. *Information Processing Letters*, 93(3):149–153, 2005.

[V'y97] Vladimir V. V'yugin. Effective convergence in probability and an ergodic theorem for individual random sequences. *SIAM Theory of Probability and Its Applications*, 42(1):39–50, 1997.

[Wal82] Peter Walters. *An Introduction to Ergodic Theory*, volume 79 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1982.

[Wei93] Klaus Weihrauch. Computability on computable metric spaces. *Theoretical Computer Science*, 113(2):191–210, 1993.

[Whi93] Homer S. White. Algorithmic complexity of points in dynamical systems. *Ergodic Theory and Dynamical Systems*, 13:807–830, 1993.

[You02] Lai-Sang Young. What are srb measures, and which dynamical systems have them? *J. Stat. Phys.*, 108:733–754, 2002.

Dipartimento di Matematica Applicata, Universita di Pisa
E-mail address: `s.galatolo@docenti.ing.unipi.it`

LIENS, Ecole Normale Supérieure, Paris
E-mail address: `hoyrup@di.ens.fr`

LIENS, Ecole Normale Supérieure and CREA, Ecole Polytechnique, Paris
E-mail address: `rojas@di.ens.fr`